COVER SHEET

Book Title
Book Editors
IOS Press, 2003

Associate A User's Goal: Exhaustivity and Specificity Information Retrieval Using Ontology

Xiaohui Tao
x.tao@qut.edu.au

Abstract. In information retrieval it is difficult to extract the accurate information to satisfy a user's information need. Based on the goals, we categorize the searches into two groups: information search and navigational search, and proposed a method using ontology to extract the the specific or general context for the given query and perform the search using it. An IR system using the method can be more efficient as it performs the search associating to the user's particular goal.

Keywords. User's goal, ontology, exhaustivity, specificity, information retrieval

1. Introduction

In information retrieval it is difficult to extract the accurate information to satisfy a user's information need. A great difficulty is that we can not read the user's mind to acquire what he/she really wants. Sometimes a user knows the exact information need, and just puts on a query which he/she thinks best represents the need and then performs a search. We name this kind of search as information search, as the user's goal is obtaining the information that the query represents to. Sometimes a user may not have an exact idea about what they want, and just puts on a query to search it out in order to undertake another precise search, or a user may just simply want to get access to an online resource [1,3,5]. We name this as navigational search, as the users' final goal is not the search results but just an intermediator in the information retrieval process. If we know a user's goal, we may be able to serve better results to the user.

We argue that this can be achieved by using ontology to extract the different levels (specificity or exhaustivity) of context for a given query to retrieve based on a user's goal. The results indicate that while performing a specificity search for the goal of information search the results are with better precision, whereas while performing an exhaustivity search for navigational search the results are with better recall.

2. Method

A well known feature of ontology is that a node on the upper bound contains more general concept, and covers broader semantic area than nodes on the lower bound with more

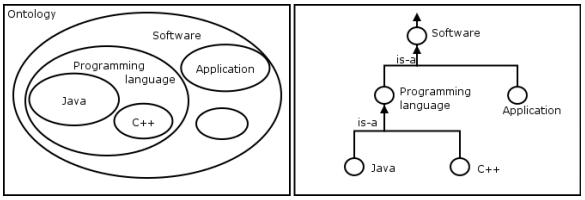

Figure 1. Example of Ontology

specific concept [2,4,6]. Figure 1 illustrates the feature using a simple ontology. A specific concept of "Java" holds an *is-a* relation to the general concept of "Programming language", and the same as "Programming language" to "Software". By common sense we know that "Programming language" contains multiple languages rather than just "Java". And except for "Programming language" "Application" holds *is-a* to "Software" as well. If search using concept of "Programming language", we may have results covering concepts of "Java", "C++", etc. However, if search using just the concept of "Java", we will have results covering only about "Java" but not "C++", because the semantic area is restricted by a more specific concept.

Based on this feature, we may perform search using the given query's context extracted from ontology depending on a user's goal. If a user wants an information search, we can perform a **specificity retrieval**, which is using the query's context extracted from the concepts on the lower bound of ontology. For example, for a query of "programming language as java", we can use the context extracted from the specific concept of "Java". Because the user wants an information search and has already had an exact idea about the information need, we are supposed to serve the results as precise as possible. "Java" is more specific than "Programming language", and so that it can be with more precise results. Whereas, if the user wants a navigational search (e.g. the user wants another programming language which is like java, but he/she can not recall the name of it and may just wants to search it out, on the above example query), we can perform an **exhaustivity retrieval**, which is using the given query's context extracted from the concepts on the upper bound of ontology. This time we prefer the general concept of "Programming language" rather than "Java", because "Programming language" covers broader semantic area, and the user will have more exhaustive results relevant to "programming language" to remind his/her mind.

The specificity and exhaustivity information retrieval have different focuses, and end with different levels of the precision and recall rate. A specificity retrieval uses more specific context to perform search, its results are with better precision, but trading with recall. An exhaustivity retrieval uses more general context to perform search, its results are with better recall, but some precision may be scarified.

3. conclusion

We categorize a user's searching goals into two groups: information search and navigational search, associate the goal with a method using ontology to extract the specific or general context for the given query, and then present the user different search results based on the goal. An Information Retrieval system using this can be more efficient as the information retrieval associates to a user's particular need.

References

[1] A. Broder: A taxonomy of web search, *SIGIR Forum, ACM Press*, **36** (2002), 3–10.

[2] J. Gonzalo, F. Verdejo, I. Chugur, & J. Cigarran: Indexing with WordNet synsets can improve text retrieval, ACL/COLING Workshop on Usage of WordNet for Natural Language Processing, 1998.

[3] U, Lee, Z. Liu, & J. Cho: Automatic identification of user goals in Web search, In Proceedings of the 14th international conference on World Wide Web, *ACM Press*, (2005), 391–400.

[4] G. A. Miller: WordNet: a lexical database for English. *Commun. ACM, ACM Press*, (1995), **38**, 39–41.

[5] D. E. Rose, & D. Levinson: Understanding user goals in web search, In Proceedings of the 13th international conference on World Wide Web, *ACM Press*, (2004), 13–19.

[6] K. M. Sim: Web agents with a three-stage information filtering approach, In Proceedings of International Conference on Cyberworlds 2003, (2003), 266–273.